

Computerunterstützte Inhaltsanalyse in der Massenkommunikationsforschung: ein Überblick über deutschsprachige Forschungsarbeiten

Schönbach, Klaus

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Schönbach, K. (1979). Computerunterstützte Inhaltsanalyse in der Massenkommunikationsforschung: ein Überblick über deutschsprachige Forschungsarbeiten. In R. Mackensen, & F. Sagebiel (Hrsg.), *Soziologische Analysen: Referate aus den Veranstaltungen der Sektionen der Deutschen Gesellschaft für Soziologie und der ad-hoc-Gruppen beim 19. Deutschen Soziologentag (Berlin, 17.-20. April 1979)* (S. 768-776). Berlin: Deutsche Gesellschaft für Soziologie (DGS). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-135704>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Computerunterstützte Inhaltsanalyse in der Massenkommunikationsforschung - Ein Überblick über deutschsprachige Forschungsarbeiten

Klaus Schönbach

Massenkommunikationsforschung und elektronische Inhaltsanalyse - eine Verbindung, die eine Fülle praktischer Erfahrungen erwarten läßt. Inhaltsanalyse als Methode systematischer Beschreibung von Kommunikationsaussagen ist - nach Noelle-Neumann - "maßgeblich im Zuge der Erforschung der Massenmedien entwickelt worden".¹ Darüber hinaus ist dem Untersuchungsmaterial der Massenkommunikationsforschung gleichsam ex definitione oft ein Massencharakter eigen: Weniger das einzelne Dokument als die Vielzahl massenhaft verbreiteter Kommunikationsinhalte steht im Vordergrund der Analyse. Beide Charakteristika prädestinieren die Massenkommunikationsforschung geradezu, bei der Kategorisierung großer Textmengen den Computer mit heranzuziehen. In Wirklichkeit jedoch ist die Verwendung computerunterstützter Verfahren in der Publizistikwissenschaft des deutschsprachigen Raums mehr als selten. Die wenigen Inhaltsanalysen aus dem Bereich der Massenkommunikationsforschung, die bisher mit Hilfe elektronischer Kategorisierung angefertigt wurden, stammen bezeichnenderweise aus Institutionen, die sich eher am Rande ihres eigentlichen Aufgabengebiets mit Massenmedien befassen - aus dem Seminar für Sozialwissenschaften an der Universität Hamburg und aus dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim. An beiden Orten hat sich die Forschung dabei auf eine wichtige Frage der Publizistikwissenschaft weitgehend konzentriert: Welches Bild der Realität entwerfen Massenmedien in ihren Aussagen?

In Hamburg hat Alexander Deichsel dazu die Aufmacherüberschriften von zehn Tageszeitungen im Verlauf eines Jahres herangezogen.²

Es geht ihm in seiner Untersuchung um die "Quantifizierung von Aufmerksamkeitsorientierung".³ Kategoriensystem für diese Beschreibung ist das HKW, das Hamburger kommunikationssoziologische Wörterbuch. Deichsel entwickelt es aus seinen Schlagzeilen und lehnt

sich dabei stark an das "Harvard III Psychosociological Dictionary"⁴ an. Das HKW versucht, die psychischen und sozialen Charakteristika des Schlagzeileninhalts so umfassend wie möglich zu beschreiben und ist dadurch für eine Vielzahl von Fragestellungen verwendbar: Welches Personal bevölkert die Zeitungsschlagzeilen? Welche Rollentypen werden dabei geschildert? Welche materiellen Objekte kommen vor? Aber auch: Finden in der Welt der Zeitungsüberschriften Konflikte oder Kooperationen statt? usw. Mit dem HKW kann Deichsel eine Reihe von "Nachrichtenfaktoren" empirisch nachweisen, die bis zu diesem Zeitpunkt oft nur vermutet worden waren.⁵ Nachrichtenfaktoren sind Merkmale eines tatsächlichen oder künstlichen Ereignisses, die dieses Ereignis für den Journalisten berichtenswert machen. Dazu gehören laut Deichsels Studie z.B. die folgenden Charakteristika: Zeitungsschlagzeilen sind handlungsentensiv, hoch personalisiert, schildern negative Entwicklungen und beschränken ihre Berichterstattung weitgehend auf geographisch und kulturell nahe Ereignisse.⁶

Deichsel und eine Hamburger Arbeitsgruppe verfolgen seit kurzem diesen Ansatz weiter. In einem von der Deutschen Forschungsgemeinschaft geförderten Projekt wollen sie ein neues Kategoriensystem zur Beschreibung öffentlicher Aufmerksamkeitsorientierung in Tageszeitungen validieren. Diese Studie ist der Rest des sogenannten "headline"-Projekts, das solche Aufmerksamkeitsstrukturen langfristig und international vergleichend überprüfen wollte.⁷

Mitarbeiter des Zentrums für Umfragen, Methoden und Analysen (ZUMA) in Mannheim und ich haben in einer dritten hier vorzustellenden Studie einen ähnlichen Forschungsansatz verfolgt wie Deichsel in Hamburg - mit einem Unterschied: Gleichberechtigt neben der Frage, welchen "Realitätsvorschlag" die Medien dem Rezipienten überhaupt präsentieren, steht ein weiteres Forschungsziel: Was veranlaßt die Journalisten als Kommunikatoren, gerade dieses - wie wir wissen⁸ - in vielen Punkten vorhersagbare Bild der Wirklichkeit zu übermitteln? Dazu liegt inzwischen eines der größten maschinenlesbaren Textkorpora aus dem Bereich der Massenkommunikationsforschung vor. Es umfaßt mehr als 166.000 Wortformen. Sein Inhalt: die Lokalteile dreier Tageszeitungen (Mannheimer Morgen,

Rhein-Neckar-Zeitung, Rheinpfalz), soweit sie über Mannheim berichten. Vom 6. bis zum 18. Juni 1977 wurden dazu insgesamt 28 solcher Lokalteile gesammelt und danach zunächst konventionell verschlüsselt. Das dabei verwendete Kategoriensystem ist aus der Tradition der Nachrichtenwert-Forschung entstanden und lehnt sich stark an das von Winfried Schulz entwickelte und in seinem Buch "Die Konstruktion von Realität in den Nachrichtenmedien" beschriebene Klassifikationsschema an.⁹ Zur Zeit wird nun der Versuch unternommen, Kategorie für Kategorie auch maschinenlesbar zu operationalisieren und das gespeicherte Textkorpus elektronisch zu vercoden. Zur Validierung dient dabei die konventionelle Verschlüsselung. Erste Erfahrungen wurden bereits mit einem Wörterbuch gesammelt, das kriminelle Handlungen kategorisieren soll. Analog zur konventionellen Inhaltsanalyse, die den Schweregrad von in der Zeitung berichteten Straftaten erhoben hat, ist dazu ein maschinenlesbares Wörterbuch konstruiert worden.¹⁰ Eine erste Analyse des Textes hat Ergebnisse gebracht, die denen der konventionellen Vercodung stark ähneln: eine relativ seltene Erwähnung krimineller Handlungen insgesamt, aber eine deutliche Betonung von Kapitalverbrechen. Ein zweites Wörterbuch versucht, den Ereignisort der Lokalbeiträge automatisch zu verschlüsseln. Es enthält alle konkreten Ortsangaben - wie Straßennamen, öffentliche Gebäude, Plätze usw.¹¹ An diesem operationalisierten Kategoriensystem zeigt sich einer der wichtigsten Vorteile der elektronischen Inhaltsanalyse: Sie ist in der Lage, die Vielzahl von Ereignisorten ungleich feiner zu differenzieren, als man das von einem menschlichen Vercoder ohne Gefahr für die Verlässlichkeit verlangen dürfte. Hier ist automatisierte Verschlüsselung dabei, die konventionelle Textanalyse nicht nur zu ersetzen, sondern sie sogar zu verbessern.

Ein zweites Beispiel für diese Entwicklung ist eine Studie von Hans-Jürgen Hippler und Albrecht Kutteroff. Die beiden Autoren haben die Kategorien "Prominenz" und "persönlicher Einfluß" am Mannheimer Korpus maschinenlesbar operationalisiert.¹² Sie sind dabei einen entscheidenden Schritt über die konventionelle Kategorisierung hinausgegangen: Sie hatte die Prominenz von am Ereignis beteiligten Personen einfach auf einer Vierer-Skala nach dem Grad ihrer Bekanntheit verschlüsselt, "persönlichen Einfluß" nach dem

Ausmaß ihrer Macht. Hippler und Kutteroff ziehen den Computer als Hilfsmittel zur Präzisierung dieser Informationen heran, ermitteln auch die organisatorische Herkunft der Handlungsträger und stufen ihre Funktion ein. Ihre Befunde geben Aufschluß über die Abbildung lokaler Machtstrukturen in der Presse. Eines der wichtigsten Ergebnisse: Wenig artikulationsfähige Randgruppen bedürfen der Prominenz bzw. des persönlichen Einflusses anderer am Ereignis beteiligter Personen, um in die Zeitung zu kommen.

Die fünfte und letzte Studie, die hier zu beschreiben ist, ist die einzige, die eine grundsätzlich andere Fragestellung behandelt als die bisher vorgestellten. Sie analysiert nicht die Realitätskonstruktion von Massenmedien, sondern die Komplexität von Zeitungstexten und den vermutlichen Reiz, sie zu lesen. Je komplizierter ein Text formuliert ist und je seltener ein Leser sich von ihm angesprochen fühlen kann, umso geringer ist die Chance, daß dieser Text die Schranke der selektiven Wahrnehmung durchbricht - das ist die Ausgangshypothese der Verfasser.¹³ Um diese Frage zu untersuchen, haben die Autoren die vier 1975 in Hamburg erscheinenden Tageszeitungen herangezogen: die beiden Kaufzeitungen Bild und Hamburger Morgenpost und die Abonnementblätter Die Welt und Hamburger Abendblatt. Eine relativ kleine Textstichprobe (etwa 8.800 Wortformen) aus den einzelnen Sparten dieser Zeitungen wurde in Einzelwörter zerlegt und in zwei Kategoriensysteme eingeordnet. In einem sind Indikatoren versammelt, mit denen die Verfasser Komplexität von Zeitungssprache messen wollen. Dabei geht es um die Verwendung des Passivs, von Konjunktionen, Präpositionen, um Satzlänge, Zahl der Nebensätze usw. Das zweite Wörterbuch versucht, Indikatoren für die Distanz zwischen dem Textinhalt und seinem Leser zu finden. Hier vermuten die Autoren z.B. Identifikationsmöglichkeiten für den Rezipienten, wenn ihm viele Personen vorgestellt werden, die der gleichen sozialen Schicht angehören wie der Leser selbst. Darüber hinaus aber sind im Distanzwörterbuch Kategorien enthalten, von denen die Arbeitsgruppe Aufmerksamkeitssteigerungen beim Publikum erwartet, wie z.B. emotional befrachtetes Vokabular, die Schilderung von Sensationen oder von angsterregenden Vorgängen. Die Ergebnisse dieser Studie zeigen, daß Abonnementzeitungen eine deutlich kompliziertere Sprache

verwenden als Kaufzeitungen - sicher kein unerwarteter Befund, aber doch der Nachweis, daß sich elektronische Inhaltsanalyse auch für diese und ähnliche Feststellungen sinnvoll einsetzen läßt.¹⁴

Vier der fünf vorliegenden Forschungsarbeiten aus der Bundesrepublik Deutschland ist offensichtlich gemeinsam, daß sie im wesentlichen eine bestimmte Untersuchungsfrage der Massenkommunikationsforschung herausgegriffen haben: Welche Züge trägt die mediale Wirklichkeit? Das ist eigentlich auffallend - denn: Sicher handelt es sich hier um eine der wichtigsten Fragen, die die Publizistikwissenschaft in den letzten Jahren bewegt haben, aber es gibt natürlich auch noch eine Reihe anderer Probleme in der Massenkommunikationswissenschaft, die ebenso brisant und aktuell wären - so z.B. die Untersuchung von Gewaltdarstellungen in den Medien, die "Ausgewogenheit" von Hörfunk und Fernsehen, Sozialisation durch Massenkommunikation usw. Die Beschränkung der spärlichen computerunterstützten Medienuntersuchungen auf ein Thema läßt sich sicher nicht allein inhaltlich erklären, sondern hängt offenbar auch mit den technischen Besonderheiten elektronischer Inhaltsanalyse zusammen: Die Entscheidung für oder gegen den Einsatz automatisierter Verfahren stellt sich dem Inhaltsanalytiker ja meist als Kosten-Nutzen-Rechnung dar. In diese Kalkulation gehen einerseits die Präzision, die absolute Verlässlichkeit und die Schnelligkeit elektronischer Inhaltsanalyse ein, aber auch der Zeit- und Kostenaufwand für die Übertragung des Materials in maschinenlesbares Format und für die Konstruktion von Wörterbüchern. Die Entscheidung für die elektronische Inhaltsanalyse fällt nun immer dann besonders leicht,

1. wenn das zu untersuchende Material ohnehin nicht auf dem - in der Regel nicht maschinenlesbaren - Originaldatenträger belassen werden kann, sondern - etwa vom Tonband - abgeschrieben werden muß. Dann kann man diese Abschrift mit wenig zusätzlichem Aufwand auch maschinenlesbar vornehmen.¹⁵ Gerade das aber ist in der Massenkommunikationsforschung bisher eigentlich nur bei der Analyse von Hörfunk- und Fernsehsendungen bzw. bei der Auswertung von Interview-Protokollen sinnvoll gewesen. Zeitungs- und Zeitschriftenbeiträge - das am meisten inhaltsanalytierte Material - nochmals

abzuschreiben, wird oft als zu mühsam empfunden.

2. Zeit und Kosten reduzieren sich mit Hilfe elektronischer Verfahren auch dann, wenn sicher ist, daß das Untersuchungsmaterial nicht nur einmal kategorisiert wird, sondern mehrfach - etwa unter verschiedenen Fragestellungen: ein in der Massenkommunikationsforschung bisher recht seltener Vorgang.

3. Ein dritter Grund, sich für oder gegen den Einsatz computerunterstützter Verfahren zu entscheiden, ist die Struktur des Kategoriensystems. Der Computer ist dann besonders leistungsfähig gegenüber einem menschlichen Vercoder, wenn er mit einem umfangreichen Kategoriensystem arbeitet, das zugleich nur geringe intellektuelle Leistungen des Forschers bei der operationalen Definition der Codes, der Konstruktion des Wörterbuchs also, erfordert. Bei kleinen, aber komplexen Kategorienschemata herrscht die Tendenz vor, sie lieber einem menschlichen Vercoder zu erklären und im übrigen auf seine sprachliche und wissenschaftliche Kompetenz zu vertrauen, als "forschungshygienisch" (Alexander Deichsel) vorzugehen und alle Indikatoren für eine Kategorie exakt anzugeben.

Vielleicht hat gerade dieser dritte Posten in der Kosten-Nutzen-Kalkulation vor dem Einsatz computerunterstützter Verfahren dazu geführt, sich so stark auf die Beschreibung der Medienrealität mittels Nachrichtenfaktoren zu konzentrieren. Denn hier handelt es sich in der Regel um einfach zu definierende Indikatoren, von denen allerdings eine Fülle im Text gefunden werden muß: Wer ist am öffentlichen Geschehen beteiligt? Welche Themen werden angesprochen? Wo finden Ereignisse statt? usw.

Allerdings haben sich alle genannten Studien nicht nur dieser Beschränkung unterzogen, sondern auch im Umfang des analysierten Materials (im Gegensatz zu vielen konventionellen Inhaltsanalysen) ihren Fallstudiencharakter noch nicht verloren. Dafür ist wohl vornehmlich der erste in die Kosten-Nutzen-Rechnung eingehende Faktor verantwortlich - die notwendige Überführung von Aussagen in maschinenlesbares Format. Diesen engen Horizont versucht eine Arbeitsgruppe am Institut für Publizistik in Münster aufzubrechen. Sie geht von der Überlegung aus, daß so gut wie alle deutschen Tageszeitun-

gen und Publikumszeitschriften lochstreifen- oder magnetbandgesteuert gesetzt werden. Dadurch entsteht gleichsam als Abfall eine maschinenlesbare Kopie des Zeitungsinhalts. Es ist technisch trivial, diese Daten in ein Format zu übersetzen, das eine Bearbeitung mit den vorhandenen Inhaltsanalysesystemen möglich macht. Das Rechenzentrum der Universität Münster hat ein ähnliches Problem beispielsweise schon für den maschinenlesbaren Text von Stern und Zeit gelöst, der vom Computer in Blindenschrift übersetzt wird.

Die Münsteraner Arbeitsgruppe folgt in ihrem Projekt einem Beispiel aus den USA, das Carroll DeWeese beschrieben hat:¹⁶ Der Satz der "Detroit News" wird tagtäglich über Magnetband in ein elektronisches Inhaltsanalysesystem eingegeben und vollautomatisch auf die in ihm vorhandenen Themen überprüft. Der Forschungsansatz dieser Analyse dort wie in Münster ist die sogenannte "agenda-setting"-Theorie: Rundfunk und Presse formulieren die Themen, die man unter bestimmten Bedingungen kurze Zeit später als Diskussionsstoff in der Bevölkerung wiederfinden kann.¹⁷ Die Überprüfung dieser Theorie ist in der Vergangenheit oft daran gescheitert, daß immense Datenberge dabei zu untersuchen und zu vergleichen sind. Hier kann elektronische Inhaltsanalyse den Aufwand entscheidend vermindern.

Aber auch ohne die Möglichkeit, sogenannte "direct copies" von Medienaussagen zu analysieren, wäre die Massenkommunikationsforschung in der Lage, eine Vielzahl von Untersuchungen fast kostenfrei durchzuführen, wenn mehr als bisher Texte und Wörterbücher ausgetauscht würden.

Warum die deutsche Publizistikwissenschaft sich bisher so wenig um elektronische Inhaltsanalyse gekümmert hat (die geringe Anzahl computerunterstützter Studien ist ja offensichtlich nicht auf vielfältig gescheiterte Versuche zurückzuführen), hat sicher über die genannten Gründe hinaus auch eine Ursache, die mit der Struktur des Faches an den deutschen Universitäten zusammenhängt. Publizistik ist in vielen Instituten traditionell hermeneutisch und historisch-qualitativ ausgerichtet.¹⁸ Erst in jüngerer Zeit werden auch empirisch-quantifizierende Untersuchungen durchgeführt. Zum zweiten ist

selbst dort, wo es schon so etwas wie ein empirie-freundliches Klima gibt, elektronische Inhaltsanalyse nicht bekannt oder mit einem Negativ-Image behaftet: Sie gilt als teure Spielerei mit wenig konkretem Nutzen. Ihre technische Entwicklung, die sie inzwischen weit über die Einwort-Analyse hinausgeführt hat, wird nicht zur Kenntnis genommen. Die Verlässlichkeit elektronischer Inhaltsanalyse, die Schnelligkeit und Exaktheit ihrer Kategorisierung auch der Massenkommunikationsforschung zugänglich zu machen, heißt zuerst, diese Barrieren zu beseitigen und Informationsdefizite zu vermindern.

ANMERKUNGEN

- 1 Elisabeth Noelle-Neumann: Methoden der Publizistikwissenschaft, in: Elisabeth Noelle-Neumann, Winfried Schulz: Publizistik. Frankfurt a.M.: Fischer 1971, S. 175-194, S. 176.
- 2 Alexander Deichsel: Elektronische Inhaltsanalyse. Berlin: Spiess 1975.
- 3 ebenda, S. 66.
- 4 Vgl. dazu: Philip J. Stone u.a.: The General Inquirer: A Computer Approach to Content Analysis. Cambridge/Mass.: The M.I.T. Press 1966, S. 170 ff.
- 5 Einen Überblick über diese Forschungsrichtung gibt Winfried Schulz: Die Konstruktion von Realität in den Nachrichtenmedien. Freiburg, München: Alber 1976, S. 11 ff.
- 6 Alexander Deichsel: a.a.O., S. 108 ff.
- 7 Vgl. dazu: Alexander Deichsel: Mass Communication: The Headline Project, in: Alexander Deichsel, Knut Holzschek (Hrsg.): Maschinelle Inhaltsanalyse. Materialien: 1. Universität Hamburg: Seminar für Sozialwissenschaften 1976 (als Manuskript vervielf.) S. 78-89
- 8 Vgl. z.B. Winfried Schulz: a.a.O., S. 115 ff.
- 9 Ergebnisse der konventionellen Inhaltsanalyse finden sich in Klaus Schönbach: Die isolierte Welt des Lokalen, in: Rundfunk und Fernsehen (26) 1978, S. 260-277.
- 10 Vgl. dazu auch: Klaus Schönbach: Nachrichtenwerte und computerunterstützte Inhaltsanalyse, in: ZUMANACHRICHTEN 2, Mai 1978, S. 3-11, und: Peter Ph. Mohler, Klaus Schönbach: Deutsche Diktionäre für computerunterstützte Inhaltsanalyse, Version 2/1979. Mannheim: ZUMA 1979 (als Manuskript vervielf.), S. 52.
- 11 Vgl. auch: Peter Ph. Mohler, Klaus Schönbach: a.a.O., S. 32.
- 12 Hans-Jürgen Hippler, Albrecht Kutteroff: Kommunale Machtstruktur und Lokalpresse. Vortrag auf dem 19. Soziologentag, 17.-20. April 1979, Berlin.
- 13 Seminar für Sozialwissenschaften, Universität Hamburg: Komplexität und Distanz. Hamburg: Seminar für Sozialwissenschaften 1976 (als Manuskript vervielf.).
- 14 Die Verschlüsselung mit dem Distanzwörterbuch bringt heterogene Ergebnisse, was die Autoren auf inkonsistente Kategoriendefinitionen zurückführen.
- 15 Dafür stehen z.B. sogenannte OCR-Schriften zur Verfügung, die ein automatisches Lesegerät identifizieren und in digitale Impulse umwandeln kann.

- 16 L.Carroll DeWeese III: Computer Content Analysis of Printed Media: A Limited Feasibility Study, in: Public Opinion Quarterly (40) 1976, S. 92-100; und: Ders.: Computer Content Analysis of "Day-Old" Newspapers. A Feasibility Study, in: Public Opinion Quarterly (41) 1977, S. 91-94.
- 17 Vgl. dazu z.B. : Maxwell McCombs, Donald Shaw: The Agenda Setting Function of the Mass Media, in: Public Opinion Quarterly (36) 1972, S. 176-187.
- 18 Nur 10 Prozent der dem Informationszentrum für sozialwissenschaftliche Forschung gemeldeten Inhaltsanalysen stammen aus dem Bereich der Kommunikationswissenschaft und Publizistik (Zahlen von 1977, zusammengestellt von Ekkehard Mochmann).